Recorded Future

Anatomy of Cyber Threats, Vulnerabilities, and Attacks

ACTIONABLE THREAT INTELLIGENCE FROM ONTOLOGY-BASED ANALYTICS

ABOUT RECORDED FUTURE

The open web is both a platform to create attacks and a source of information to prevent attacks. To shift the balance of power in your favor, our revolutionary technology organizes the public web for analysis to provide you future, present, and past insight for emerging cyber threats.

Our Temporal Analytics™ Engine structures data around cyber security events, actors, locations, and time to give you forecasting power. Operating at a massive scale in real time, Recorded Future scans, collects, and analyzes hundreds of thousands of web sources in seven languages, and processes billions of events to cast the widest open source intelligence net and deliver tailored, timely insights to you.

Recorded Future organizes the web for analysis of past and future cyber security events, which enables analysts to generate meaningful threat intelligence to more accurately and proactively defend their organization.

To enable this, unstructured text from internet forums, websites, paste sites, news articles, blogs, tweets, etc. is transformed into structured information, which can be visualized for human analysis, aggregated to support (algorithmic) quantitative analysis, and analyzed to detect anomalies and trends. The end goal is to forecast future events and even create automated predictive models. To ensure threat intelligence is accurate and quickly actionable, it's critical that it's based on a standardized ontology to ensure a consistent integration with security products and other intelligence sources, and enable confusion-free collaboration with analyst teams.

This white paper introduces the data model that underpins Recorded Future's real-time threat intelligence solution. It describes what entities are involved in representing cyber threats, vulnerabilities, and attacks, how these entities are related in our cyber ontology, and how cyber events represent relationships between different involved entities.

We're not alone in trying to structure the complex world of cyber security. According to MITRE, STIX™ is "a collaborative community-driven attempt to define and develop a standardized language to represent structured cyber threat information. The STIX Language intends to convey the full range of potential cyber threat information and strives to be fully expressive, flexible, extensible, automatable, and as human-readable as possible."

We always strive to follow standards when possible. We have monitored the growing adoption of standards like STIX for

representing threat intelligence information. These standards are a useful point of reference for explaining our approach. In fact, there's a straightforward mapping between STIX and much of our entity ontology and CyberAttack and CyberExploit events, as described in this paper.

Our approach is designed for the full breadth of threat information that's found on the web, which includes reporting from defenders, security researchers, and more. This consists of ambiguously reported clues about current threats, and even reporting from threat actors, such as statements linking attacks to hacktivist "operations" and hashtags. We've designed our approach with the expressiveness needed for these additional characteristics.

Creating Structure With Entities and Events

Internal data regarding cyber security tends to be highly structured (e.g. in the form of network logs, data from intrusion detection systems, etc.). External, open source intelligence comes primarily in the form of unstructured text, and needs to be organized before it can be quantitatively analyzed and correlated with internal data. Recorded Future uses natural language processing (NLP) to extract entities and events from unstructured text, and organizes (mostly) static relationships between entities into ontologies to relate them to each other.

Entities

Recorded Future uses entities to model concrete and abstract "nouns," such as persons, organizations, companies, products, and technologies.

An entity represents a physical or virtual object, and can have several

names associated with it. We refer to the alternative names for the same entity as synonyms. For example, the malware entity “DDoS” has several synonyms, including “Distributed Denial of Service.” Synonyms allow the user to not have to worry about alternative names and spellings of an entity when querying the Recorded Future system.

For the cyber domain, we have introduced a number of domain-specific entity types.

Malware

This entity type is used to represent malware. The name is intended to be the “non-technical name” or “street name” used to discuss the malware. Examples include Zeus and Duqu.

Malware entities are detected by harvesting industry expert and government sources as well as through a statistical entity extractor.

MalwareSignature

A malware signature is the “technical name” of a malware, as reported by a cyber security company, for example Trojan.W32.Zeus.

MalwareSignature entities are identified by regular expression detectors.

CyberVulnerability

A vulnerability is a bug or a weakness that can be directly used by a hacker to gain access to a system or network. A CyberVulnerability entity represents a vulnerability defined, for example, in the US National Vulnerability Database (NVD), such as CVE-2014-0094.

CyberVulnerability entities are harvested through a set of regular expression detectors, defined for the different companies and organizations that assign vulnerability identifiers.

MalwareCategory

The MalwareCategory entity type is used to group Malware entities into logical groups, for example by what kind of systems they target (e.g. POS malware and Android malware).

AttackVector

Malware uses different attack vectors to gain access to a computer in order to deliver a payload or malicious outcome. These classes of attack vectors is represented by the AttackVector entity type. Examples include SQL Injection and Phishing.

Operation

This entity type represents operations, typically defined by hacktivist organizations. Examples include OpIsrael and OpGCHQ.

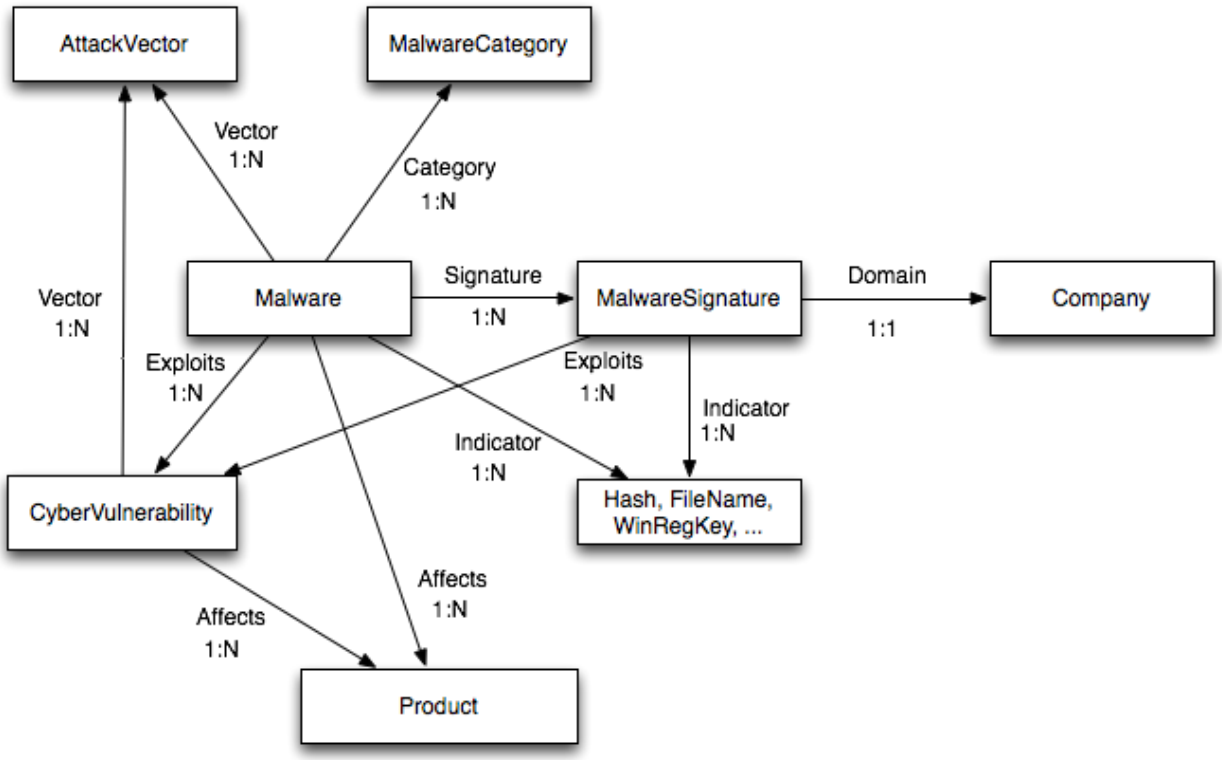
Operation entities are defined by a regular expression entity detector.

Ontology

The Recorded Future cyber ontology represents primarily static relationships between entities, as described in the picture below, and the following example.

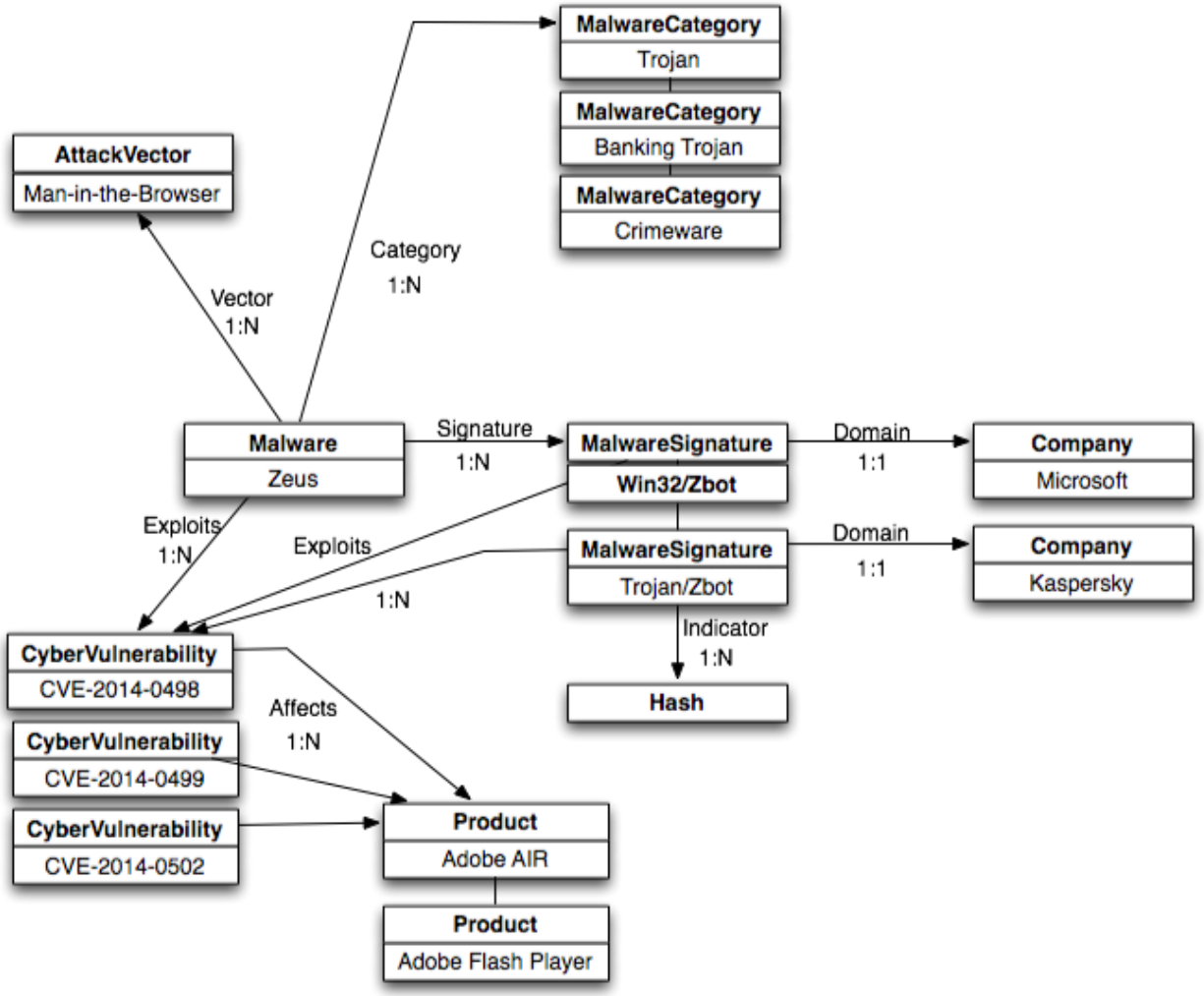
The central Malware entity type can for example be associated

with an arbitrary number of AttackVector, MalwareCategory, MalwareSignature, CyberVulnerability, and Product entities, as well as with an arbitrary number of technical indicators such as hashes, file names, Windows registry keys, etc.



Example

Below is an example of (parts of) the ontology information for the Zeus malware:



Events

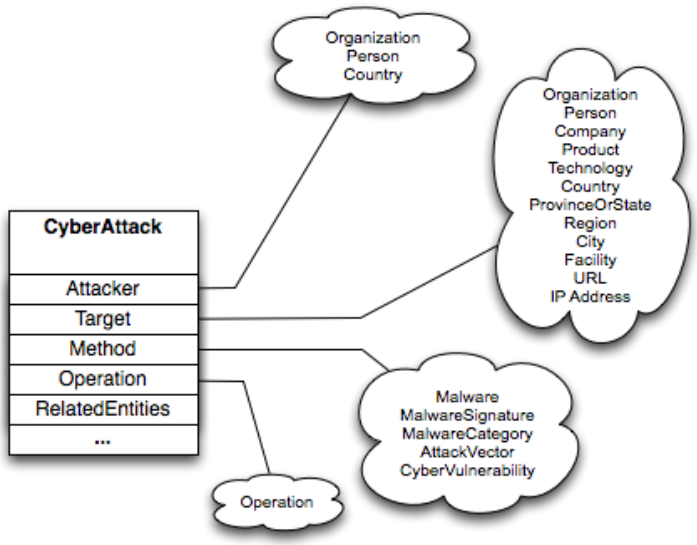
Events in Recorded Future capture dynamic relationships between entities, and are also associated with an event time. Currently there are two event types specific to the cyber security domain: CyberAttack and CyberExploit.

Just like synonyms allow us to abstract away from alternative spellings and names for the same entity, events allow us to abstract away from the exact wording used to describe an event. For example, phrases such as “XYZ Bank was hacked,” “The hackers went after XYZ Bank,” and “Data breach hits XYZ Bank” all result in a CyberAttack event where

XYZ Bank is designated as target.

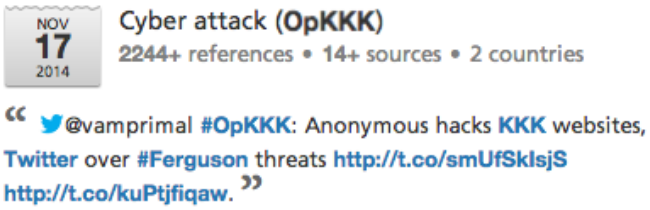
CyberAttack

CyberAttack is the event type used to represent information about a cyber attack. Note that information might be partial, for example only a Target or an Attacker and a Method might be known. In general, a CyberAttack event is described by an (optional) Attacker, Target, and Method attribute. In some cases, a (hacktivist) Operation attribute can also be identified. Entities which occur in a sentence discussing a cyber attack but that cannot be assigned a specific role get collected in the RelatedEntities attribute:

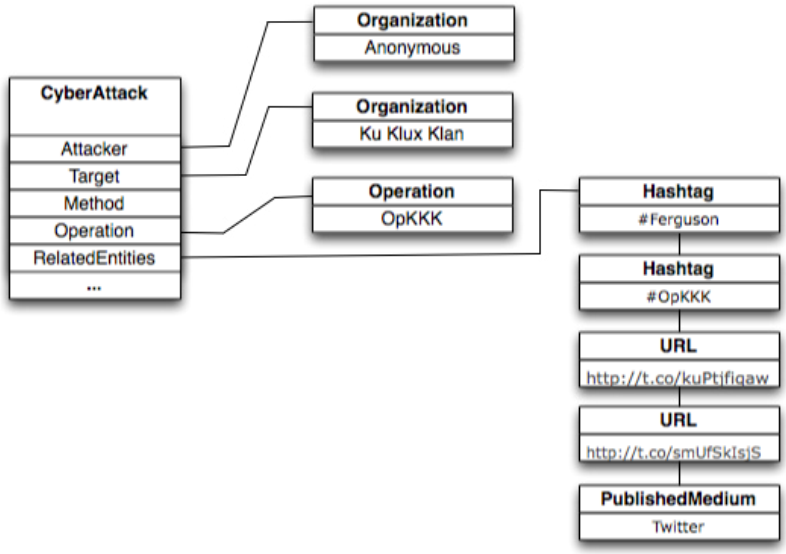


Example

The following is an example of a cyber attack where several attributes have been identified:



This results in the following event data structure. Note that “KKK” gets identified as a synonym and is automatically resolved to the entity “Ku Klux Klan.”

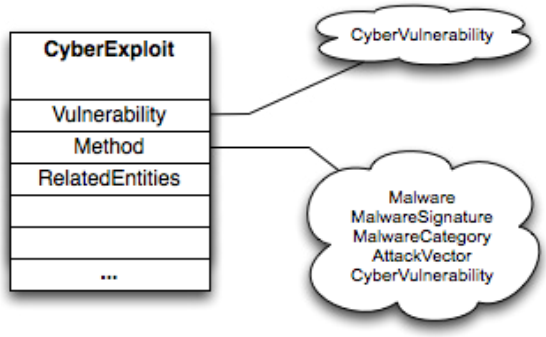


CyberExploit

CyberExploit is the event type used to represent when a known vulnerability (e.g. one which has been assigned some CyberVulnerability identifier) has been exploited, either malignantly (in the wild) or as a Proof of Concept (PoC) to illustrate its potential.

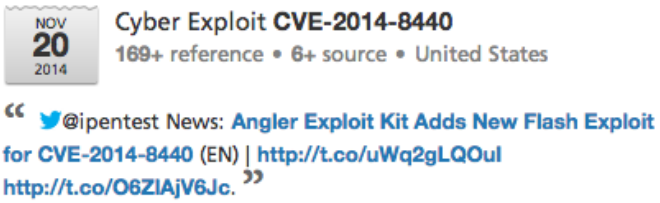
The CyberExploit event currently only relates the CyberVulnerability to the method used to exploit it. Future versions might add specific products or other targets hit by the exploit. Entities which occur in a sentence discussing a cyber exploit but that cannot be assigned a

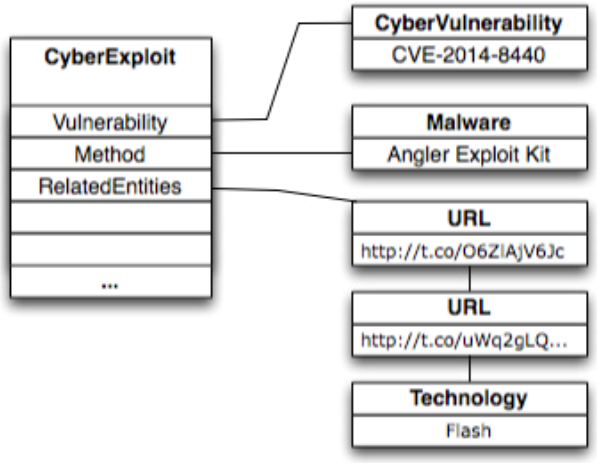
specific role get collected in the RelatedEntities attribute:



Example

The following is an example of a CyberExploit event connecting CVE-2014-8440 to the "Angler Exploit Kit" malware:



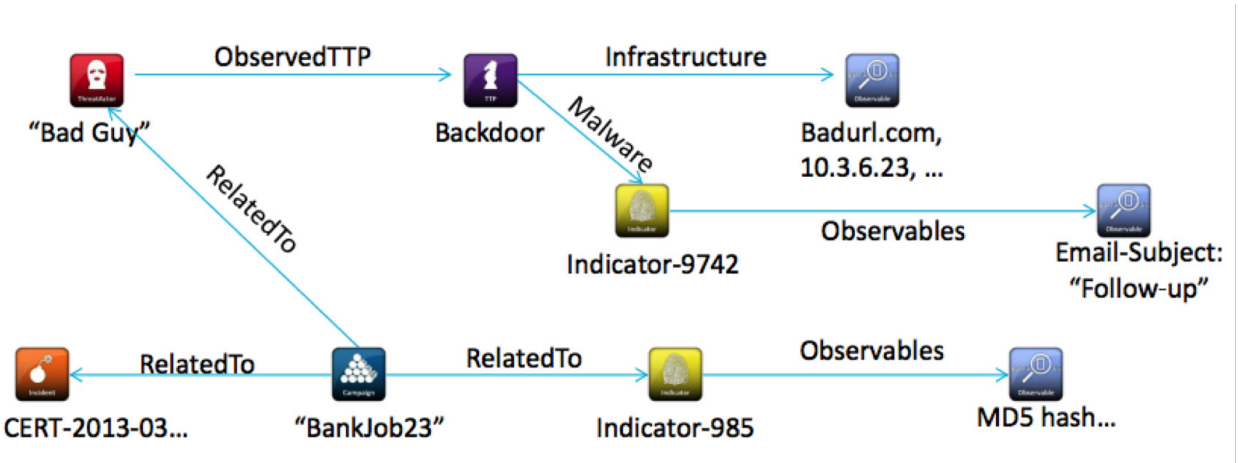


Mapping to STIX

As mentioned above, our representation of cyber entities and events can be mapped into the emerging STIX standard:

Structured Threat Information eXpression (STIX) v1.1 Architecture (Source)

As a simple example, consider the following STIX data:

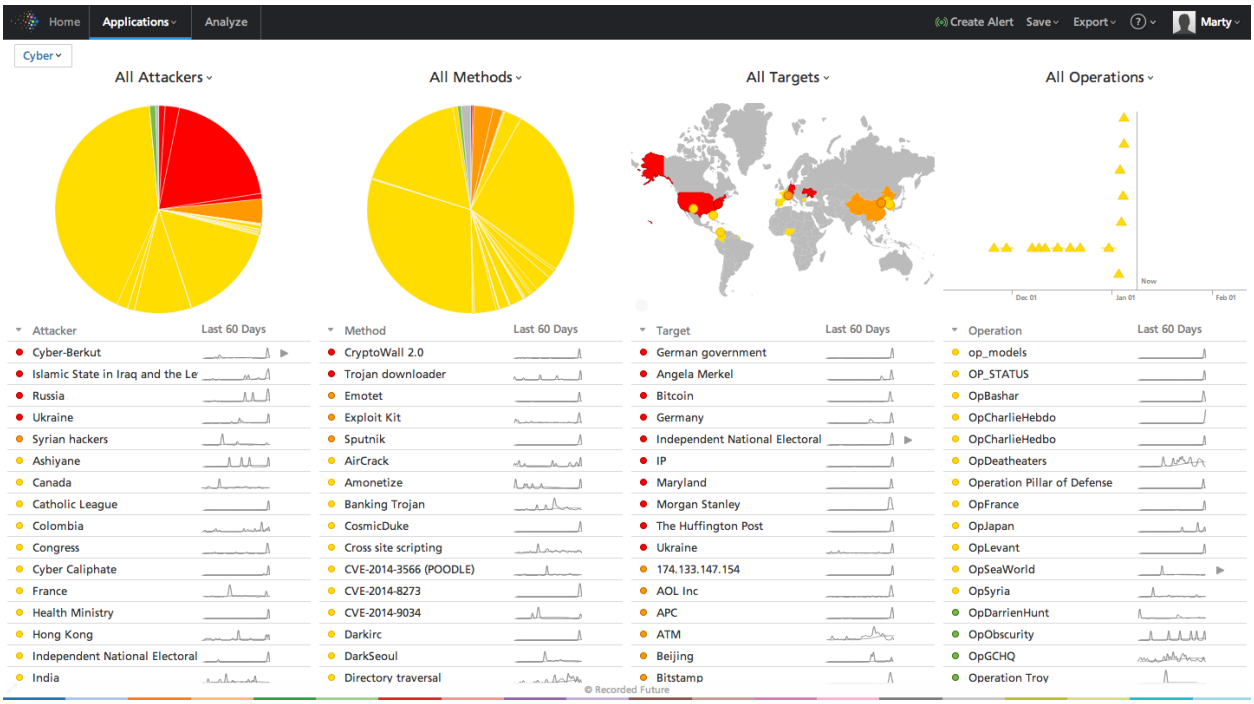


Nodes in this graph correspond to entities in our system, and edges represent a mix of ontological information and events (e.g. a CyberAttack event with “Bad Guy” as attacker and “Backdoor” as method, and potentially “BankJob123” as an operation). The exact mapping between our representation and STIX is however beyond the scope of this paper.

Conclusions

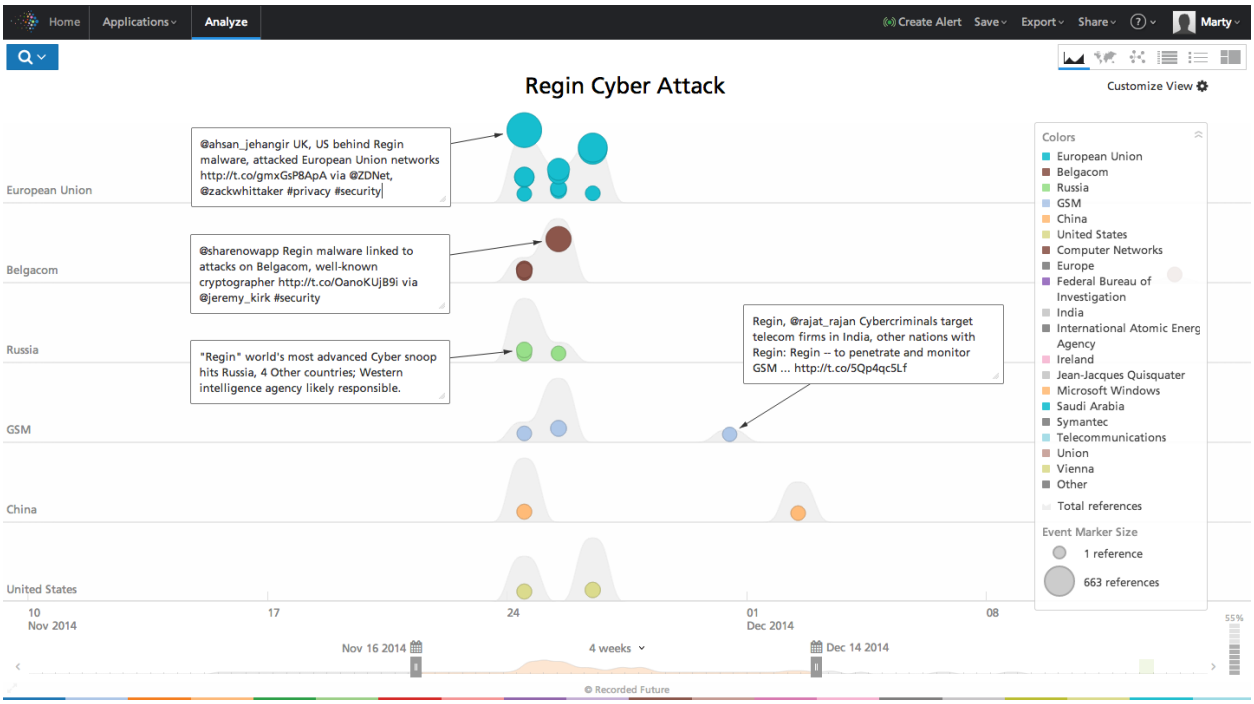
Once information has been structured into entities and events, it can be aggregated, clustered, and used for different kinds of analyses.

As an example, the Recorded Future Cyber product shows a real-time view of the most important threat actors, targets, methods and operations, all based on data structured as we have described in this white paper.



Recorded Future Cyber

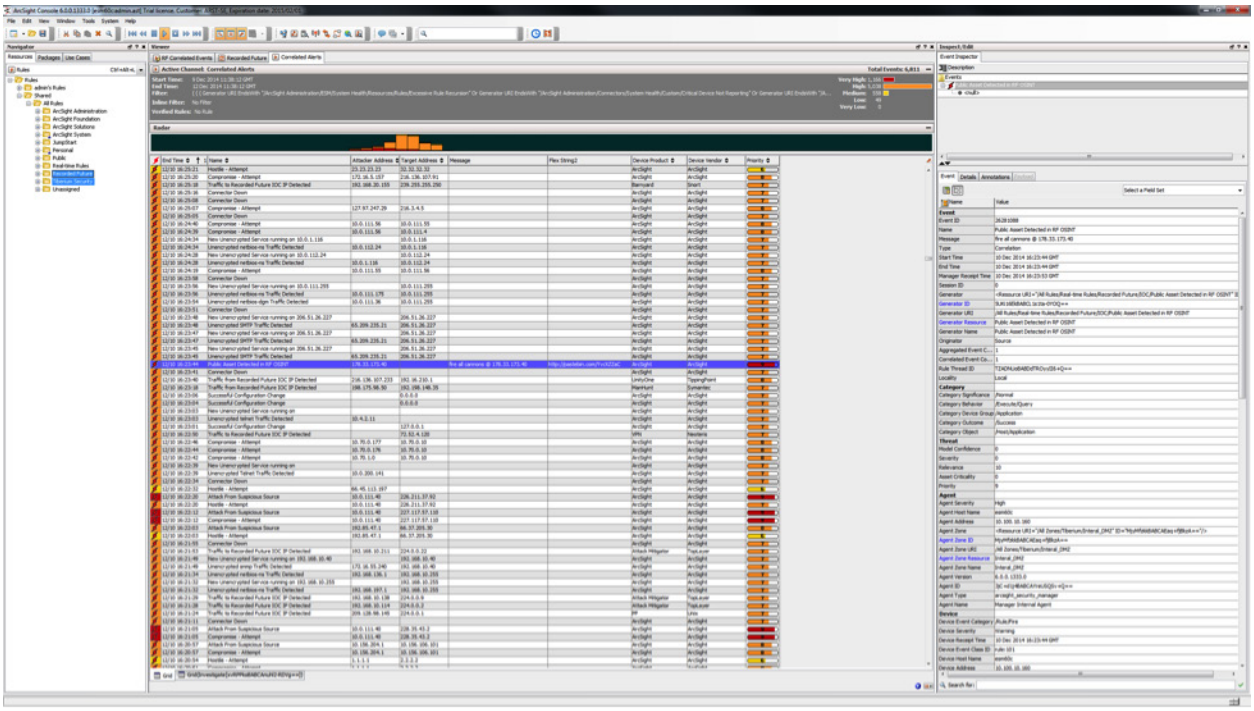
From this top level view an analyst can drill down, for example, to look for targets affected by a certain malware, such as Regin:



Recorded Future Enterprise

The structured representation and the ability to search for a specific event type, and with certain attribute values, is part of what makes Recorded Future the best choice for creating a more insightful world.

Analysis does not have to be confined to the Recorded Future system. Through our integration with the HP ArcSight security information and event management (SIEM) solution, security operations center (SOC) analysts can directly link to Recorded Future's real-time threat intelligence solution for actionable insight on relevant technical indicators.



Recorded Future Integration via the HP ArcSight SIEM Solution

--

References

CVE: http://cve.mitre.org/
NVD: http://nvd.nist.gov
STIX: http://stix.mitre.org/
OpenIOC: http://openioc.org/

